

平成 29 年度の調査・分析活動報告

1. 調査・分析業務の内容と目的

独立行政法人日本学術振興会グローバル学術情報センター（以下、「情報センター」と略）は、学振の諸事業に関する情報の収集・蓄積と、これらを用いた調査・分析を主な業務としているが、その主なものは以下のとおりである。

- (1) 科学研究費助成事業（科研費）等の改善に向けた業務
- (2) 学振諸事業の支援の効果のエビデンスを明らかにするための業務
- (3) 諸外国の学術研究動向の情報の収集と活用に向けた業務

これらの業務については、情報センター所長の下、3 名の分析研究員と、その業務を支援する 3 名の分析調査員が取り組んでいるが、中でも「(1) 科学研究費助成事業等の改善に向けた業務」については、2 名の分析研究員が情報センターに設置されたデータベースを用いて分析を行った。

2. 具体的な調査・分析課題

平成 29 年度に行った調査・分析の主な内容は以下のとおりである。なお、これらの業務のうち、各分析研究員による活動の報告は 2 ページ以降に掲載している。

(1) 科学研究費助成事業等の改善に向けた業務

① 項目反応理論 (IRT) に基づく評点の標準化に関する調査研究

情報センターでは、科研費の審査における評点のデータを分析することにより、今後の審査方針の検討に対し有効な知見を提供するための検討を行っている。その一つの手法が書面審査の評点から推定される潜在的なスコア θ の計算方法である、IRT に基づく標準化である。平成 29 年度においては、潜在的な評点 θ の多次元化についての調査研究が行われた。この分析結果については、3. の「A. 項目反応理論 (IRT) に基づく評点の標準化に関する調査研究」において報告されている。

② 主成分分析と項目反応理論 (IRT) 2 次元パラメータモデルの対比

科研費の書面審査の評点のデータを用いた分析としては、前年度に引き続き科研費審査データ

を対象とした主成分分析も行った。前年度は審査員の付した評点を基に審査結果の検証を目的として行った分析に加え、上記の IRT の有効性についても分析も試みた。平成 29 年度においてはこの IRT の有効性に関する分析を更に進め、IRT の 2 次元パラメータモデルとの対比を目的とした分析を行った。この分析結果については、3. の「B. 主成分分析と項目反応理論 (IRT) 2 次元パラメータモデルの対比」において報告されている。

③ 審査員候補の統計的推薦システムの開発

情報センターが行う科研費等の事業の実施の改善に向けた他の取組としては、審査員候補者の推薦がある。学術システム研究センターにおいて行われる審査員の選定は、個々の同センター研究員の専門的な知見に依存する部分が多いが、この支援のために開発を行っているものが、統計的潜在意味解析の手法により候補者を推薦するシステムである。この分析結果については、3. の「C. 審査員候補の統計的推薦システムの開発」において報告されている。

④ 情報センターが保有するデータベースの活用に向けた取組

情報センターが保有するデータベースを情報センター構成員以外の学振内部の役職員も利用し、様々な条件を設定することにより集計を行うことが出来る「科研費データベース活用システム」と名付けたツールを開発し、学術振興会内部で運用を開始した。

また、このシステムから非開示情報を削除した上で、インターネット上で公開するためのシステムの開発を行った。

(2) 学振諸事業の支援の効果のエビデンスを明らかにするための業務

学振の諸事業による支援がどのような効果を上げたかについてのエビデンスを明らかにするため、情報センターデータベースを用い、例えば科研費の基金化が支援を受けた研究の成果にどのような影響を及ぼしたかといった依頼に対応した分析を行った。

また、文献データベースを利用した分析を継続的に実施しており、学振の日常の業務において活用するとともに、分析手法の検討も行っている。その具体的な内容は、3. の「D. Scopus のデータを用いた分析の実施と利用手法の検討」において報告されている。

(3) 諸外国の学術研究動向の情報の収集と活用に向けた業務

諸外国の学術研究動向の調査を継続的に実施しており、平成 29 年度は米国、ドイツ、英国等を中心に各国の動向をより深く掘り下げた調査を行った。この調査の目的は学振の国際交流事業をはじめとする事業の改善に向けた検討に資することであり、これら諸外国の学術研究動向に関連した業務の内容は、3. の「E. 海外学術研究動向の調査」において報告されている。

3. 分析研究員による調査・分析活動報告

以下は、3 名の分析研究員による平成 29 年度の調査・分析活動の報告である。

A. 項目反応理論 (IRT) に基づく評点の標準化に関する調査研究 (持橋分析研究員)

IRT とは、計量心理学の分野で開発された、テスト理論に基づく得点の精密な統計モデルであ

り、学術振興会の場合、各申請に対してその良さを示す潜在的な得点 θ を推定する。 θ は 0 を中心に標準正規分布（ほぼ $-4 < \theta < 4$ 程度）で分布しており、図 1 のように θ が大きいほど、5 段階で 4 や 5 などの高評価が得られる確率が高い。

具体的には、IRT の一種である段階反応モデル (Samejima 1969) において、スコア θ の申請が審査員 j から評点 c を得る確率は以下で与えられる。この確率を各 c についてプロットしたものが図 1 となっている。

$$p(y=c|\theta, j) = \sigma(a_j(\theta - b_{jc})) - \sigma(a_j(\theta - b_{j,c+1}))$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

IRT において、得点の確率は審査員 j ごとに異なることに注意されたい。IRT では、観測された得点をもとに各審査員の潜在的な評価パラメータ a_j, b_{jc} を最適化によって求めることができる。

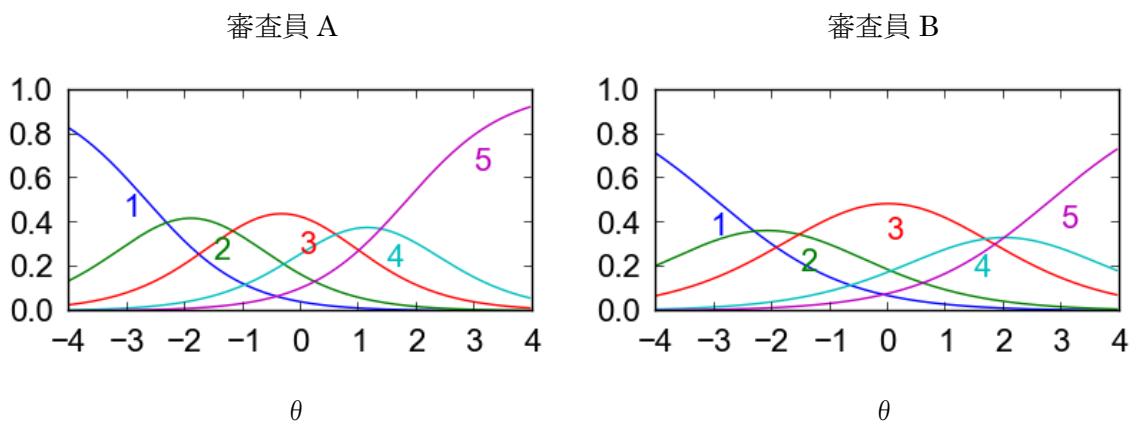


図 1. IRT による科研費申請に対するスコアの応答曲線の例。審査員によって各曲線の形状が異なり、そのパラメータはデータから統計的に学習される。

これに対して、現在使われている T スコアの変換式は以下のものである。

$$y' = 3 + \frac{y - \mu_j}{\sigma_j} \times 0.6$$

T スコアでは上の式に基づき、スコア y を標準スコア y' に形式的に変換する。

ここで μ_j, σ_j は審査員 j のスコアの平均および分散を表す。IRT と異なり、T スコアは形式的なスコアの変換であるため、多数の同点が生じる。また数学的には、T スコアは

- (1) 各審査員の審査の「確度」 a_j をモデル化していない
- (2) 各審査員がスコア 1,2,3,...,5 ごとに持っている閾値をモデル化していない
- (3) 各申請のスコア θ を 2 次元以上にすることができない

といった問題を持っている。しかし IRT では、同点になる確率は 0 である。

こうした背景から、IRT に基づく得点の標準化についてこれまで検討してきた。

この時、資料「B. 主成分分析と項目反応理論 (IRT) 2 次元パラメータモデルの対比」にあるように、実際には申請について複数の評価軸がある、すなわち、 θ は 2 次元以上にとれる可能性がある。また、この計算は複雑であり、実際に学術振興会で継続的に運用するには、統計的な背

景や数学的計算を適切に理解・実行できる企業の協力を得ることが必要である。

そこで今年度は、数理に強い企業に委託し、IRT に基づいて統計的パラメータと各申請の潜在的得点 θ を計算するプログラムを開発した。これにより計算はパッケージ化され、容易に運用できるようになった。また、 θ の多次元化は特に複雑な最適化が必要となるが、これも実装され、IRT に基づいた複数の主成分を計算する基盤を構成することができた。

多次元化された IRT での推定例を図 2 に示す。これに加えて、実際には多次元の潜在空間において、図 1 のような反応曲面が審査員ごとに存在している。

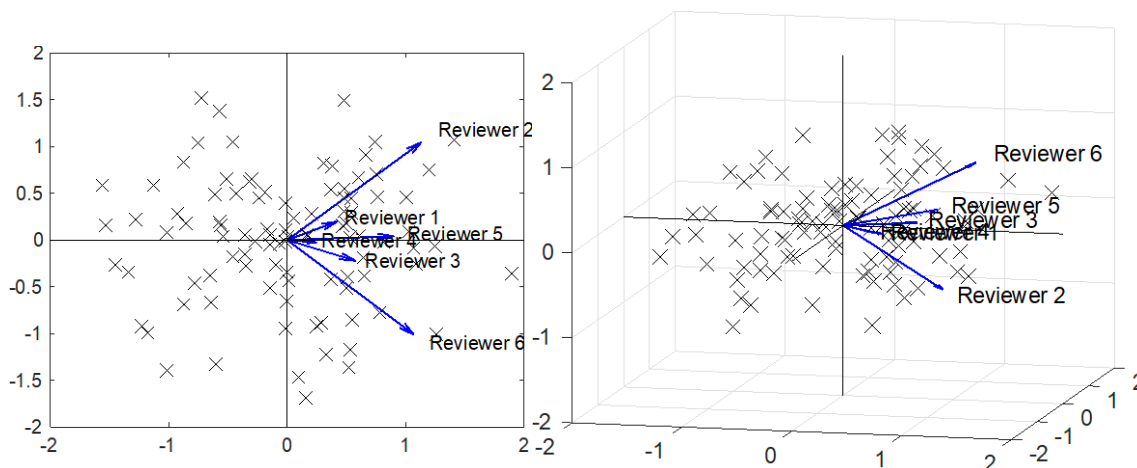


図 2. 多次元の IRT による各申請の θ (×印) と各審査員の評価ベクトル(矢印)。

左は 2 次元の場合、右は 3 次元の場合を推定結果の例を表す。全体を総合した評価は、各審査員ベクトルに沿った評価の平均、すなわち評価ベクトル自体の平均を使って求められる。

この空間は、単位ベクトル $(1,0,\dots,0)$ が評価ベクトルの期待値となるよう、角度の確率分布を用いて統計的に基準化されている。

B. 主成分分析と項目反応理論(IRT)2次元パラメータモデルの対比 (三浦分析研究員)

科研費申請審査データに対して、昨年度は 1 次元パラメータの IRT モデルの分析結果と主成分分析の対比を行った。そこで分かったことは、第 1 主成分の寄与率が高い審査セットにおいては、IRT モデルから推定される θ の値と第 1 主成分スコアの相関が強く、また同様に T スコアとの相関も強いことである。これは、各申請の良さを測る評価軸が審査員チームの中でほぼ 1 本であり、第 1 主成分と θ がそれぞれ申請の質の評価を表していると解釈してよいことを示す。

第 1 主成分寄与率が高くない審査セットにおいても、それが第 2 主成分の寄与率とある程度の差があれば、上記のことが成立する。しかし、第 1 と第 2 の主成分寄与率が拮抗する審査セットにおいてはそれが成立しない。これに対して、今年度は IRT のパラメータ θ が 2 次元である場合の計算プログラムが完成したので、それによる分析結果と対比させて評点分析手法の検討を行った。

主成分の強さを表す寄与率の大きさを用いて審査セットを寄与率が高位、中位、低位の 3 通りに分けて、推定された項目反応理論 2 次元パラメータ (θ_1 , θ_2) と主成分スコアの関係調べた。

その結果、大まかではあるが以下のような特徴がみられた。

(1) 寄与率高位グループは θ_1 が第 1 主成分スコアと相関が強く、 θ_2 の数値がほぼゼロである。
 (2) 寄与率中位グループは第 1 主成分の寄与率が中程度であっても、第 2 主成分寄与率との差がある程度大きければ（多くの場合がこれに該当するのだが）、第 1 主成分スコアと θ_1 の相関が高く θ_2 は主成分スコアと強い関係を持たない。
 (3) 上記のように高位と中位のグループについては項目反応理論モデルのパラメータを 1 次元にとどめておいても申請の良さを図る一本の評価軸を作ることがほぼ可能であるが、低位グループについては一括りの結論が得られず、主成分と θ の関係は多様であることがわかった。これまでに扱った審査セットに関しては、以下の 3 つのタイプの事例がみられた。

1 つ目のタイプは、第 1 と第 2 の主成分寄与率が 0.3 以下のレベルで極めて拮抗しており、 θ_1 が第 2 主成分スコアと、そして θ_2 が第 1 主成分スコアと相関が強いことがみられた。 θ_1 と θ_2 はゼロから離れた数値でありどちらも無視できない。つまり、拮抗する評価軸が 2 本存在することを示している。これは θ_1 、 θ_2 がそれぞれ一つだけの主成分と相関が強く他の主成分とは相関がかなり弱い事例である。主成分と θ の対応する番号が逆転しているが、これは 1 次元 θ モデルでは θ と第 1 主成分スコアの相関がかなり低く第 2 主成分スコアとの相関が高かったために（図 3）、主成分の性質に基づいて θ を解釈することが難しかった事例である。 θ を 2 次元にすることにより主成分との関係が見やすくなった。

2 つ目のタイプとしては、 θ_1 と θ_2 がそれぞれ無視できない値を持ち、しかし 1 本の主成分だけに強い相関を持つわけではなく 2 本あるいは 3 本の主成分とある程度の相関を持つ、言い換えればこれらの数本の主成分の一次結合でほぼ近似される関係にある。

3 つ目のタイプは、2 本あるいは 3 本以上の主成分寄与率が 0.15 以上 0.3 未満程度の低いレベルにあり、2 次元 θ の θ_2 はほぼゼロに近い数値であり無視出来て（図 4）、 θ_1 は第 1 主成分とは相関が弱い第 2、第 3、さらに第 4 などの主成分とは相関がある程度強い。言い換えれば θ_1 は第 2 と第 3 などの主成分スコアの一次結合でほぼ近似される関係にある。

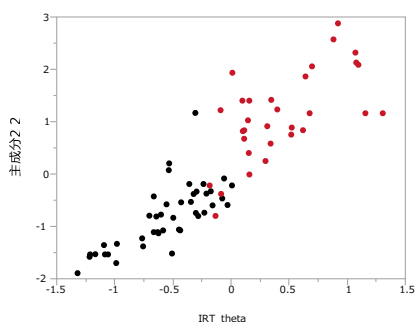


図 3. 1 次元 θ と第 2 主成分の相関事例

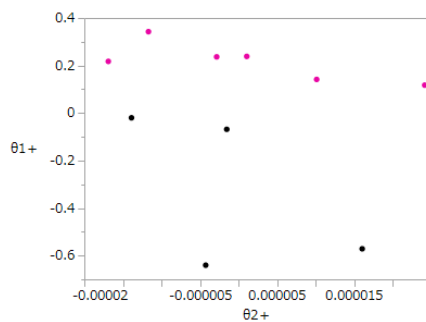


図 4. θ_1 の数値がゼロに近い事例

C. 審査員候補の統計的推薦システムの開発（持橋分析研究員）

科研費の審査にあたり、各区分において適切な審査員の選定が重要な問題となっている。選定を行う学術システム研究センターの専門研究員が想像できる候補の数には限界があり、特に科研費の中区分化を踏まえて、「広い範囲を審査できる審査員であるか」について客観的な指標を提供できるシステムがあれば有用だと考えられる。

(1) そこで、統計的潜在意味解析、特に LDA (Latent Dirichlet Allocation; Blei+ 2003) とよばれ

る方法をもとに、審査員の候補者の書いた論文あるいは申請書等から、審査員としてふさわしいかの推薦スコアを出力し、審査員の選定を補助するシステムについて検討を行った。

LDA では、各論文にその潜在的な「意味」を表す確率分布 θ があると仮定する。

θ はたとえば、(政治, 芸術, スポーツ, 家庭) = (0.2, 0.4, 0.1, 0.3) のような、和が 1 となる確率分布である。ここで「政治」や「芸術」といったトピックは便宜的に示したもので、実際にはすべて自動的に学習される。(※この θ は IRT での θ と意味が異なっている。)

(2) データとして、振興会の保有する科研費の申請書自体から審査員の専門を推定し、推薦することに取り組んだ。科研費申請書のある細目の全体集合を、可視化のため 3 次元のトピック分布によって表したものを図 5 に示した。この分布全体を被覆するように、審査員を選定することが目標となる。

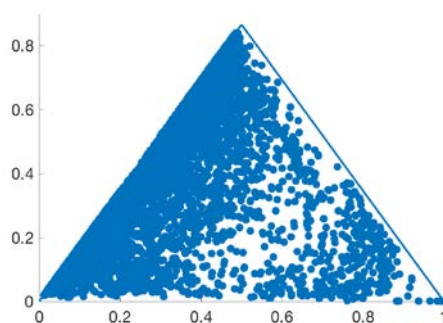


図 5. LDA (確率的潜在意味解析) によるある細目全体の潜在的内容の分布。可視化のために 3 次元に圧縮している。各点が、一つの申請書を表している。

(3) このとき、各研究者=審査員候補の書いた論文の集合は図の点集合の一部となる。その例を図 5 に示した。適切な審査員の集合を選ぶことにより、図 5 のような分野全体を被覆できるようにする、と問題を言い換えることができる。

(4) 点の集合を選んだとき、単体上でその分布は図 7 に示したようなディリクレ分布とよばれる確率分布でモデル化できる。よって、候補者の論文をすべて合わせた集合から算出されたディリクレ分布と、元の全体のディリクレ分布との距離を計算すればよい。このために Kullback-Leibler ダイバージェンスとよばれる確率分布間の標準的な距離を用いると、距離は次式で解析的に算出できる。

$$D(\text{Dir}(\boldsymbol{\alpha}) \parallel \text{Dir}(\boldsymbol{\beta})) = \log \Gamma\left(\sum_k \alpha_k\right) - \log \Gamma\left(\sum_k \beta_k\right) + \sum_k (\log \Gamma(\beta_k) - \log \Gamma(\alpha_k)) \\ + \sum_k (\alpha_k - \beta_k) [\Psi(\alpha_k) - \Psi(\beta_k)]$$

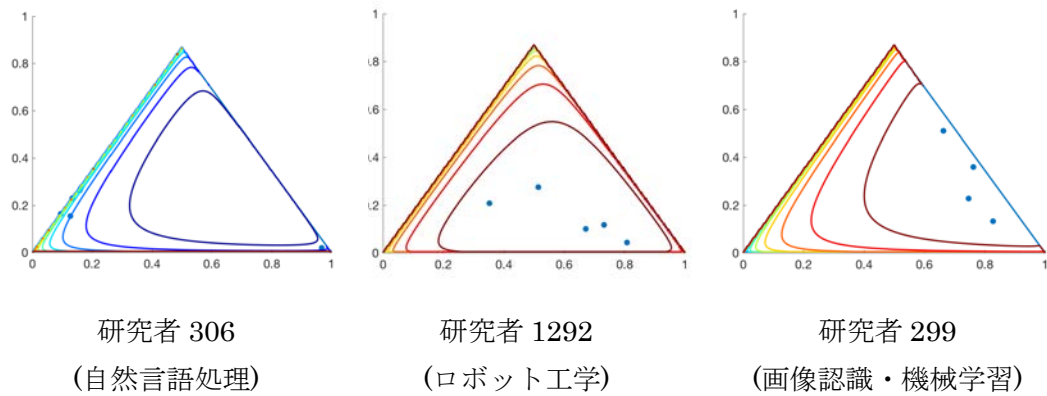


図 6. 各審査員候補の論文の内容とそれを表すディリクレ分布。

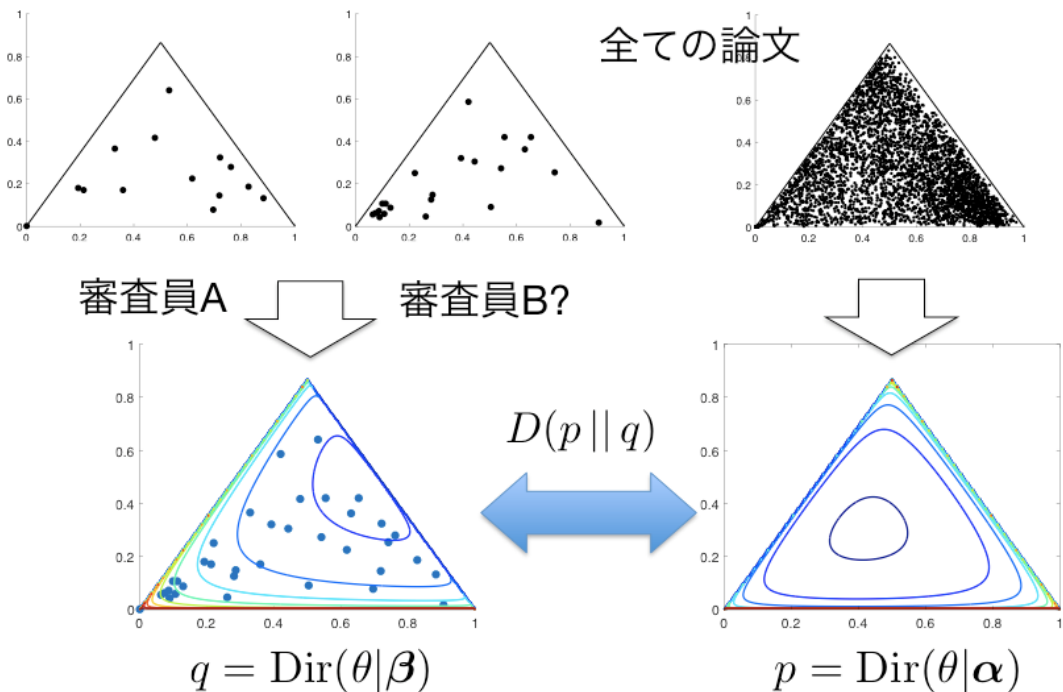


図 7. 審査員推薦スコアの計算法の概要。審査員の書いた論文の集合が、論文全てを被覆する割合を潜在空間における確率分布間の KL ダイバージェンスとして計算する。

(5) 審査員推薦システムの実装

現在の実装では、図 8 のように審査員（ここでは 306 番の研究者）を選ぶと、それを補完するような審査員の番号とスコアを順に出力する。スコア（被覆誤差）は、この審査員を含めたときの潜在トピックの分布と、データ集合全体のトピック分布の間の誤差を情報量で表す Kullback-Leibler ダイバージェンスである。

直感的には、「この人を審査員として追加した際に、全体でカバーすべき内容からどのくらい誤差が生じるか」をスコアが表している。最終的には GUI を実装して使いやすくする予定にしている。

```

>> help recommend
      recommend (alpha,gammas,authors,[selected])
      $Id: recommend.m,v 1.2 2017/09/26 09:51:36 daichi Exp $
% 推薦スコアを出力
>> score = recommend(alphas,'./model/aiinf.model.K20.gamma','./data/authors.dat',[306]);
1292    0.3442
299     0.3936
801     0.3996
916     0.4035
285     0.4036
176     0.4222
465     0.4266
1016    0.4272
939     0.4393

```

図 8. 審査員の統計的推薦システムのプロトタイプ。

(6) 審査員選定補助の仮想例

初期状態、および研究者 306, 1292 を順に選択した場合の推薦スコアは例えば表 1 の通りであった。審査員を追加するにつれ、全体の分布との誤差が減少している様子がわかる。また、図 6 の研究者ごとの被覆領域から分かるように、研究者 306（自然言語処理）を選択後は研究者 1292（ロボット工学）や 299（画像認識・機械学習）のような補完的な研究者が自動的に推薦されている。なお、306、1292 を選択後に表 1-C で推薦されている研究者はさらに補完的な分野が専門の研究者であった。

研究者 ID	被覆誤差
1060	0.9420
290	1.0952
45	1.2346
459	1.2575
306	1.2851
547	1.3892
708	1.4046

表 1-A.

初期状態での推薦スコア

研究者 ID	被覆誤差
1292	0.3442
299	0.3936
801	0.3996
916	0.4035
285	0.4036
176	0.4222
465	0.4266

表 1-B.

研究者 306 を選択後の推薦スコア

研究者 ID	被覆誤差
939	0.2202
1060	0.2285
1220	0.2410
313	0.2550
170	0.2622
1304	0.2658
1155	0.2709

表 1-C.

研究者 306 と 1292 を選択した後の推薦スコア

(7) 課題

本報告はまだ最初の段階であり、今後、対象となる論文の時期を限定することや、単体上でロジスティック正規分布のようなより複雑な分布を用いることで、この問題は解消できると考えている。

また、審査員の選定に関する資料は論文だけではなく、科研費の申請書類、論文の概要、また論文のタイトルおよびジャーナル名称など、あらゆる情報を原的に使うことができる。科研費

全体に関して、どのような情報を用いれば効果的に審査員候補の情報が得られるかについて、さらに検討していきたい。

D. Scopus のデータを用いた分析の実施と利用手法の検討（遠藤分析研究員）

Elsevier 社が提供する文献データベース Scopus 及び分析ツール SciVal を利用し、科研費助成事業や海外特別研究員事業を対象として分析を行った。具体的にはインターネット上の論文等のドキュメントに恒久的に付与される識別子であるデジタルオブジェクト識別子（DOI）により、科研費の成果論文を Scopus の文献情報と紐づけを行い、以下の観点から調査を行った。

- ・ 科研費成果論文の分野の構成
- ・ 論文の被引用度を指標とした支援の成果
- ・ 論文の共著データを利用した科研費による研究活動状況
- ・ 論文数上位大学と科研費配分額上位大学
- ・ 特許に引用された科研費成果論文

また、過去に Elsevier 社に委託したデータについて、科研費成果論文の ID を個々の研究課題の ID や研究者の ID と紐づけを行った。

E. 海外学術研究動向の調査（遠藤分析研究員）

米国国立科学財団（NSF）、米国国立保健研究所（NIH）、ドイツ研究振興協会（DFG）、英国工学物理科学研究会議（EPSRC）の各ファンディングエージェンシーを対象として事業実施の検証評価に関する以下の調査を実施した。

各ファンディングエージェンシーはいずれも調査分析部門を有していることから、その位置づけや役割について整理し、その上で各部門がどのようなデータを収集・管理しているかについて調査した。さらに各ファンディングエージェンシーがそれらのデータを用い、どのように実施する事業に関する検証評価活動を行い、また、その成果についてどのように利用しているかを取りまとめた。

さらに、各ファンディングエージェンシーが行うこれらの検証評価活動やその成果の利用について比較するとともに、課題とされている点についても整理した。

なお、この調査結果については学振内で共有するとともに、一部については CGSI レポート第 8 号として刊行した。

独立行政法人日本学術振興会グローバル学術情報センター

CGSI レポート 第 7 号

平成 30 年 3 月 29 日発行

独立行政法人日本学術振興会グローバル学術情報センター

〒102-0083 東京都千代田区麹町 5-3-1 麹町ビジネスセンター

電話：03-3263-1971

電子メール：cgsi@jsps.go.jp

